Google Slide Speaker Notes:

**Slide 1 – Title Page**

“Instacart: A Market Basket Analysis with Customer and Product Segmentation”

**Slide 2 – “What Is Instacart”?**

“Instacart is an American company currently operating a grocery delivery service in the United States and Canada. It allows customers to order groceries from participating retailers via the Instacart app with the shopping being done and delivered to the customers’ door by a personal shopper.”

**Slide 3 – Instacart CEO Quote**

“In a matter of a couple of weeks, we were already ahead of our end-of-year goal. A week later, we were ahead of our 2021 goals, and a few days after that, we were ahead of our 2022 goals.”

“Instacart CEO, Apoorva Mehta, is quoted here referencing the soaring Instacart activity in the early stages of COVID-19. As early as February 2020, Instacart reported noticing an increase in orders for home essentials such as toilet paper. During the subsequent lockdown, Instacart became an essential service for millions of Americans trapped at home. Even now as COVID restrictions are being lifted, analysts expect sustained growth in this field as Instacart users have discovered the ease of ordering online and now prefer it to the traditional grocery shopping experience. This prediction is unsurprising with Instacart boasting an estimated 9.6 million active users, over 500,000 delivery shoppers and in March 2021, the company being valued at $39 billion.”

**Slide 4 – Our Selected Topic**

“Researching and contemplating the ideal project subject proved to be a monumental task. Our group floated from one idea to the next until we came across an Instacart dataset on Kaggle. Upon initial overview of the dataset, it contained ample information including seven corresponding csv files with data on orders and prior orders and files containing categorization assistance like aisles and departments for available and sold products. This dataset is comprised of about 3 million grocery orders placed by around 200,000 customers. With agreement, our group moved forward with the idea of customer and product segmentation along with a market basket analysis. The goal is to predict purchasing patterns among these customers as well as gain insight on customer segmentation.”

**Slide 5 – Questions We Hope to Answer**

The customer segmentation questions our group hopes to answer with the Instacart data are:

* What is the distribution of annual Income by age?
* What is the distribution of gender for Instacart users?
* What is the age group of most frequent Instacart users?
* What is the overall distribution of income level for instacart users?
* How do male and female spending scores differ?

**Slide 6 – Machine Learning**

“Using KMeans and by clustering the customer data, we will try to successfully create four main categorization groups: Target, Standard, Careless, and Careful. These refer to the ratio of income level and the customer spending score. The Target group will be comprised of customers with an annual income and spending score both in high ranges. These could be ideal customers for Instacart as they have the means and regularly spend it. The Standard will have those with an annual income and spending score in mid-range. The Careless group will have the customers with a low annual income but a high spending score which suggests spending outside of their means. In contrast, the Careful group will contain those with a high annual income and a low spending score.”

**Slide 7 – The Exploratory Phase**

Cleaning the data:

Customer Data:

* Imported the customer dataset to pandas
* Renamed columns
* Dropped null values
* Changed ‘# of orders’ to ‘customer score’ which is what we will use a score of usage for the customers